Eur Spine J (2006) 15: 1239–1246
DOI 10.1007/s00586-005-0002-y

ORIGINAL ARTICLE

Jacobus J. Kriek
Shunmugam Govender

AO-classification of thoracic and lumbar fractures—reproducibility utilizing radiographs and clinical information

Received: 9 March 2005
Revised: 29 August 2005
Accepted: 20 September 2005
Published online: 21 December 2005

J. J. Kriek · S. Govender (✉)
Orthopaedics, University KwaZulu Natal, KwaZulu Natal, Durban, South Africa
E-mail: jkriek@saol.com
Tel.: +27312604297
Fax: +27312604518

Abstract This study was designed to assess the inter-observer reliability and intra-observer reproducibility of standard radiographic evaluation of 150 thoraco-lumbar fractures using the AO-classification. The influence of clinical information on agreement levels was also evaluated. Six observers (two junior and four senior residents) evaluated the radiographic images. The injuries were classified by each observer as either type A, B or C according to the AO-classification system and the levels of agreement were documented. After 3 months the injuries were again classified with the addition of the clinical findings of each patient and the level of agreement evaluated. The level of agreement was measured using Cohen's κ-test. The overall inter-observer agreement was rated as fair (0.291) in the first session and moderate (0.403) in the second. Intra-observer values ranged from slight (0.181) to moderate (0.488). The increased level of agreement in the second session was attributed to the value of additional clinical information, the learning curve of the junior residents and the simplicity of the classification.

Keywords Thoraco-lumbar · Spine fractures · AO-classification

Introduction

The reproducibility of fracture classification schemes has been a contentious issue. Appendicular skeleton classification systems have been shown to have a poor to moderate intra-observer reproducibility and inter-observer reliability [1, 2, 7, 9, 11, 15, 16, 18, 19]. There are relatively few studies assessing the reproducibility of thoracolumbar fracture classification systems [3, 17].

Fractures of the spine are complex injuries with the main concern being the effect of the injury on the mechanical and neurological stability of the spine. Development of new imaging technology, such as CT and MRI, has increased the understanding of these injuries and has contributed to development of more comprehensive classification systems. Widely utilized classifications for spinal trauma were developed by Holdsworth [10] and Denis [6]. The former system was based on a concept of two columns and the latter on three columns. Magerl et al. [14] studied a total of 1,445 consecutive patients over a 10-year period and proposed a comprehensive classification system based on the AO fracture classification scheme. Fractures were divided into three types (A, B and C). Each type was further subdivided into three groups with subgroups and specifications. The system was based on the pathomorphological characteristics of injuries.

Previous studies assessing the reproducibility of the AO-classification have suggested the addition of MRI to increase the level of agreement regarding the type classification [3, 17]. The rationale of these proposals was based on the importance of posterior element involvement in the AO-classification. The value of clinical findings were not assessed in these studies. All types of B and C fractures, with the exception of type B.3.1.1, are associated with significant injury to the posterior column. Clinical findings differentiating between types were detailed in the original description of the

AO-classification system. In type A fractures "posterior swelling and subcutaneous haematoma are not found ... There is only tenderness at the level of the fracture" [14].

Identifying type C injuries depends on the presence of signs of rotation. These are suggested by the mechanism of injury and specific radiographic findings as described by Magerl et al. These include "two-column injury; rotational displacement ... fractures of articular processes, usually unilateral; fractures of transverse processes; rib dislocations and/or fractures close to the spine; lateral avulsion fracture of the endplate; irregular fractures of the neural arch; and asymmetrical fractures of the vertebral body" [14]. Many of these features are readily identifiable on plain radiographs. The aims of this study were to assess the influence on reproducibility when clinical data was provided in addition to standard radiographs and to compare the level of agreement regarding the AO-type classification with studies that included CT and MRI evaluation.

Materials and methods

Antero-posterior and lateral radiographs and also clinical notes were collected over a 4-year period (2000–2003) of patients admitted to the spinal unit. Radiographs of 148 patients with 150 thoracolumbar fractures were obtained. Digital images of radiographs of each fracture were taken.

Clinical information from patient files included: age, gender, mechanism of injury, clinical examination findings, and the Frankel grading [8] of neurological deficit.

Six observers participated in the study: four senior residents with more than 2 years orthopaedic experience and two junior residents with less than 1 year. Junior residents were included in the study to assess the ease of familiarization with the AO-classification system. All participants were given copies of the original classification, which was reviewed 2 weeks prior to the first evaluation session.

The digital photographs were presented as a slide show. Each observer was asked to classify all fractures observed according to AO type (A, B or C). No additional clinical information was provided at this sitting.

The evaluation was repeated after a 3 month period. In the interim a further discussion of the classification system was held. No information regarding results of the first sitting was provided to the observers. The evaluation was repeated and each resident again classified each fracture according to the AO type. During this sitting, clinical information was provided including the mechanism of injury, findings on clinical examination of the spine, and Frankel grading of neurological deficit. The emphasis of clinical examination findings was on the presence of features suggestive of posterior element involvement. These included swelling, tenderness, bogginess, a palpable step, or translation in the coronal and/or sagittal plane.

All fractures were also classified by an experienced spinal surgeon who had access to the clinical information. In view of the level of agreement obtained between the observers and the spinal surgeon in the second session, this classification was used as the gold standard for statistical analysis of the relationship between AO type frequency, neurological deficit, and vertebral level of injury.

Statistics

Statistics were performed using SPSS/PC+ version 11.5 & Graphpad Quickcalcs (http://graphpad.com/quickcalcs/kappa1.cfm). The results of both the evaluations were compared to assess inter-observer and intra-observer agreement. The classification of each observer was also compared to that of the spinal surgeon. Cohen's κ-test was used to compare the results [4]. These values were categorized according to guidelines proposed by Landis and Koch [13] (Table 1). Cohen's κ-statistic is based on comparing the observed proportion of agreement between readings made by different observers or on different occasions, with the proportion of agreement, which would be expected by chance. When the categories are merely nominal, Cohen's simple unweighted coefficient is the only form of κ that can meaningfully be used. When the categories are ordinal, a weighted κ-statistic can be calculated to take into account partial agreement or close matches between adjacent categories [5, 12].

Results

Hundred and fifty fractures occurring in 148 patients were evaluated. The mean age of the patients was 31 years (range 17–72) with a male predominance of 2.5:1.The majority of the patients (93) were involved in motor vehicle accidents (71 passengers, 11 pedestrians, and 11 drivers).

Thirty-five patients fell from a height, 15 were injured by collapsing walls and 2 were assaulted. The remaining two patients were injured in a diving accident and a fall

Table 1 Landis and Koch (1977) proposed guidelines on levels of agreement

Kappa value	Agreement level
0.00–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.8	Substantial
0.81–1.00	Almost perfect

Table 2 Frankel grading of neurological deficit according to spinal level

	Frankel grade				
Spinal level	A	B	C	D	E
T1 – T9	18	3	2	3	12
T10 – L2	40	8	9	6	38
L3 – L5		2	1	2	6

from a bicycle. One hundred of the fractures occurred in the thoracic spine, 39 in the upper- (L1-2), and 11 in the lower (L3-5) lumbar spine. As expected, the majority (101 fractures) occurred at the thoraco-lumbar junction (T10-L2).

The neurogical deficit was most severe in the thoracic spine and thoraco-lumbar junction (Table 2).

All observers, except for observer E, classified more fractures as the severe type C fractures in the second session. Observer E classified 52 fractures as type C in both the first and second sessions. This observer was also the only observer to classify more fractures as type A in the second as compared with the first session (Table 3).

When two observers disagreed on the classification of a specific fracture, the level of agreement was rated higher if the disagreement was between either type B or C as compared with type A or C fractures. These values represent the weighted κ-values [5].

The intra-observer agreement levels were rated as poor in one, fair in four and moderate in one instance. Weighted values reflected three fair and three moderate levels of agreement. The mean κ-value was 0.334 (fair) and the mean weighted value was 0.411 (moderate) (Table 4).

Table 3 Type frequency of each observer

	Type	Session 1		Session 2	
		Frequency	Percent	Frequency	Percent
Observer A	A	87	58	45	30
	B	47	31.3	49	32.7
	C	16	10.7	56	37.3
Observer B	A	60	40	29	19.3
	B	48	32	51	34
	C	42	28	70	46.7
Observer C	A	25	16.7	25	16.7
	B	61	40.7	40	26.7
	C	64	42.7	85	56.7
Observer D	A	32	21.3	20	13.3
	B	57	38	49	32.7
	C	61	40.7	81	54
Observer E	A	36	24	56	37.3
	B	62	41.3	42	28
	C	52	34.7	52	34.7
Observer F	A	26	17.3	13	8.7
	B	40	26.7	44	29.3
	C	84	56	93	62

Table 4 Intra-observer agreement

Observer	Kappa values	Weighted kappa
A	0.181	0.228
B	0.307	0.387
C	0.368	0.45
D	0.363	0.431
E	0.298	0.399
F	0.488	0.571
Mean	0.334	0.411

Kappa and weighted κ-values of the inter-observer agreement are shown in Tables 5, 6, 7 and 8 and agreement levels are summarized in Table 9.

In the first session the slight levels of agreement involved one of the junior residents and the mean level of agreement was rated as fair for both κ (0.291) and weighted κ-(0.344) values. In the second session there were no slight agreement levels and the mean values improved to moderate levels of agreement for κ (0.403) and weighted κ-values (0.486).

When comparing the type classification of the residents to those of the spinal surgeon there was an increased level of agreement between the first and second sessions. The slight agreement level in the first session (0.141) was between one of the junior residents and the surgeon. All κ values were rated as moderate in the second session (Table 10). Twenty-eight fractures (19%) were classified as type A, 32 (21%) as type B and 90 (60%) as type C by the spinal surgeon (Table 11) . There was a statistically significant ($P = 0.002$) correlation between the severity of the neurological deficit and the AO-type classification. The incidence of neurological deficit was 36% with type A fractures, 50% with type B, and 76% with type C.

Table 5 Session 1 : inter-observer kappa values

Observer	A	B	C	D	E	F
A	****	0.206	0.206	0.234	0.193	0.145
B		****	0.278	0.274	0.308	0.293
C			****	0.413	0.327	0.316
D				****	0.396	0.443
E					****	0.337
F						****

Table 6 Session 1 : inter-observer weighted kappa values

Observer	A	B	C	D	E	F
A	****	0.264	0.197	0.246	0.225	0.145
B		****	0.342	0.371	0.409	0.348
C			****	0.463	0.395	0.385
D				****	0.473	0.49
E					****	0.403
F						****

Table 7 Session 2 : inter-observer kappa values

Observer	A	B	C	D	E	F
A	****	0.381	0.417	0.417	0.49	0.289
B		****	0.456	0.411	0.344	0.428
C			****	0.475	0.37	0.508
D				****	0.325	0.474
E					****	0.264
F						****

Table 8 Session 2 : inter-observer weighted kappa values

Observer	A	B	C	D	E	F
A	****	0.448	0.479	0.49	0.612	0.36
B		****	0.558	0.486	0.423	0.514
C			****	0.565	0.459	0.605
D				****	0.413	0.544
E					****	0.342
F						****

Some problem areas with regard to the evaluation of plain radiographs were identified. These included inadequate original trauma radiographs with respect to the quality of exposure or inclusion of all relevant levels. In these cases follow-up radiographs taken after resuscitation and stabilization of patients provided improved visualization. The most helpful features with regard to identification of type C fractures were rib involvement, transverse process fractures, asymmetric vertebral body fractures, and lateral endplate avulsion fractures as described in the original article. An addition helpful finding was the presence of a "double-line sign," which denoted rotation on the lateral radiograph (Fig. 4). Examples of AO fracture types A, B and C are provided in Figs. 1, 2, 3 and 4.

Table 9 Inter-observer agreement levels of both sessions

Reproducibility	Session 1	Session 1-weighted	Session 2	Session 2-weighted
Slight	2	2		
Fair	11	8	6	2
Moderate	2	5	9	11
Substantial				2
Almost perfect				

Table 10 Agreement levels between residents and spinal surgeon

Reproducibility	Session 1	Session 1-weighted	Session 2	Session 2-weighted
Slight	1	1		
Fair	4	2		
Moderate	1	3	6	6
Substantial				
Almost perfect				

Table 11 Type classification of fractures in each region

Level	Type A	Type B	Type C
T1 – T9	11	1	26
T10 – L2	14	30	57
L3 – L5	3	1	7

Discussion

The thoraco-lumbar junction (T10-L2) with its associated anatomical characteristics is especially prone to injury. These include the effect of the ribcage and sternum, facet joint orientation, increase in vertebral body size, and change in sagittal curve profile. This is reflected by the fact that two-thirds of fractures in our series occurred in this region. The severity of neurological injury in this region resulted in paraplegia in 56.4% of our patients. Fractures in the T1-9 region were less common, but 60.5% of these injuries resulted in paraplegia. The social, psychological, and financial costs associated with these injuries are immeasurable.

All the inter-observer agreement levels between the residents and the spinal surgeon reached moderate levels of agreement in the second session. In view of this finding the classification by the spinal surgeon was used as the standard in assessing the type frequency. Sixty percent of injuries in our series were classified as type C, 21.3% as type B, and 18.7% as type A. Of the 1,445 patients in the series reported by Magerl et al., the majority of patients sustained type A injuries (66.1%) with type C fractures accounting for 19.4% of the cases. The increase in the more severe type C fractures in our study is due to the large proportion of patients (93) involved in motor vehicle accidents. This can partially be attributed to poor roadworthiness of vehicles, inadequate driver education, and the disregard for the wearing of seatbelts.

In the AO-classification, fractures are classified in the order of increasing severity and instability. A statistically significant ($P = 0.002$) correlation was found in the study when comparing the type classification with the severity of neurological injury as assessed by the Frankel grading [8].

There was an increased inter-observer reliability between the first (0.291) and second (0.403) sessions, reaching moderate levels of agreement. The increase in agreement between the two sessions is attributed to the value of additional clinical information and the increased level of agreement between the observers who became more familiar with the system. Our values in the second session were similar to results obtained by Oner et al. utilizing CT (0.35 – slight) and MRI (0.39 – moderate) in the evaluation of thoraco-lumbar fractures of 53 patients [17]. The mean intra-observer agreement

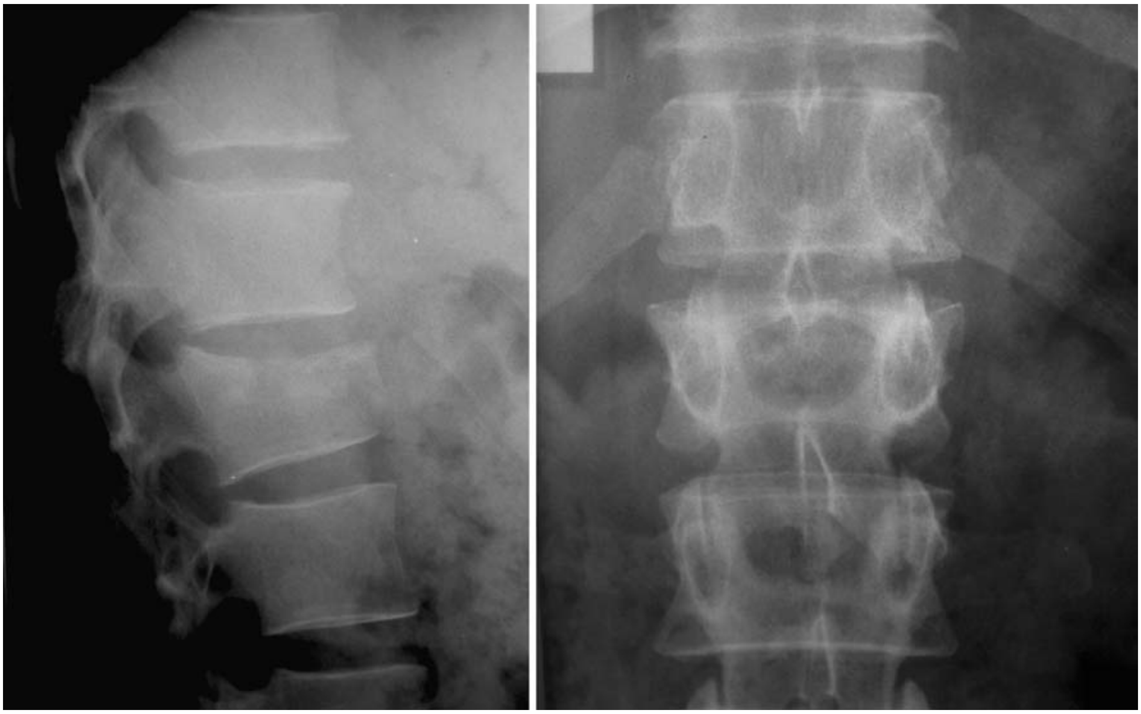

Fig. 1 A 48-year-old male patient who fell from a 3 m height and sustained an injury of the second lumbar vertebra. Frankel E with no features of posterior element involvement, clinically or on radiographs. All observers classified the fracture as an AO type A

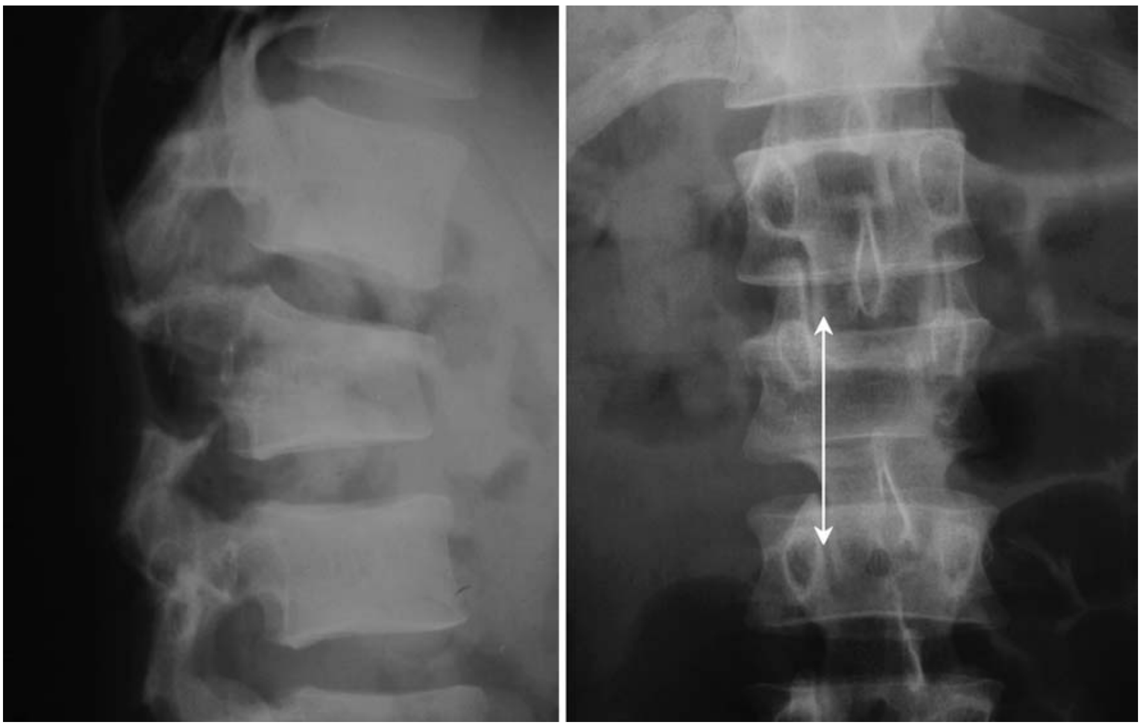

Fig. 2 A 55-year-old male driver in a motor vehicle accident. Clinical examination revealed tenderness, swelling, a haematoma and an increase in the interspinous distance. Radiographs confirm the increased interspinous distance (*arrow*) with no distinct features of a rotational component. Four of the residents and the spinal surgeon classified the fracture as an AO type B and two as an AO type C in the second session

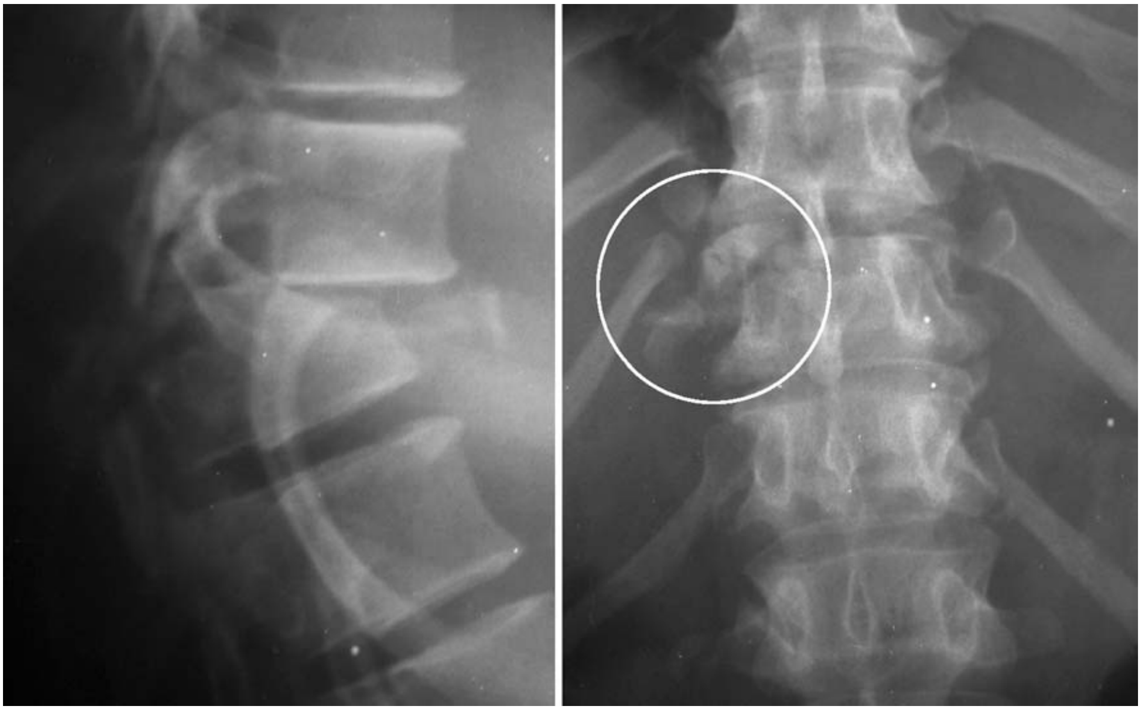

Fig. 3 This 54-year-old woman was a passenger in a motor vehicle accident. She sustained an injury at the thoraco-lumbar junction (T10-11) with Frankel A neurological deficit. Features indicative of an AO type C fracture are transverse process fractures, lateral avulsion fracture of the endplate and rib dislocations. All observers classified the fracture as a type C

in our series was fair (0.334) for the κ-values (range 0.181 – 0.488) and moderate (0.411) for the weighted values (range 0.228 – 0.571). The mean value obtained by Oner et al. for intra-observer agreement was moderate (0.41). The inter-observer agreement in the study by Blauth et al. was fair (0.33; range 0.30–0.35) when the radiographs and CT scans of 14 patients were reviewed at 22 hospitals [3].

The lowest levels of intra-observer agreement involved the two junior residents. This is attributed to the increased levels of inter-observer agreement between the junior and senior residents in the second session. We feel that this a result of the learning curve and illustrates the understanding and reproducibility of the classification.

In the AO-classification clear guidelines are provided to distinguish between the three types of fractures. Differentiation between type A and non-type A (B and C fractures) depends on disruption of the posterior ligamentous complex (the only exception being type B3.1.1). Oner et al. [17] and Blauth et al. [3] recommended the use of MRI to assist with the evaluation of posterior injury. The difficulty in using MRI is that no clear guidelines regarding MRI findings are incorporated in the original classification as stated by Oner et al. [17].

The mean levels of agreement of the AO-type classification for both the inter- and intra-observer agreement in this study were comparable with those utilizing CT and MRI in the study by Oner et al. [17]. There was some improvement in inter-observer agreement when compared with the study by Blauth et al. utilizing radiographs and CT [3]. Classification of a fracture as a specific AO-type depends in essence on the presence of posterior element involvement and whether signs of rotation are present. These features can be readily identified during clinical examination and assessment of plain radiographs. Although CT and MRI provide additional morphological information regarding the fracture personality, no increased levels of agreement have been shown with regards to the AO-type classification in aforementioned studies [3, 17]. This additional information might well be useful in the group and subgroup classification. As illustrated in the study by Oner et al. [17], where the agreement on the group classification of type A fractures was higher than the agreement on the type classification.

We have shown that additional clinical information including the mechanism of injury and clinical examination findings improved the level of agreement. These aspects were part of the original description of the AO-classification and we believe that it forms an

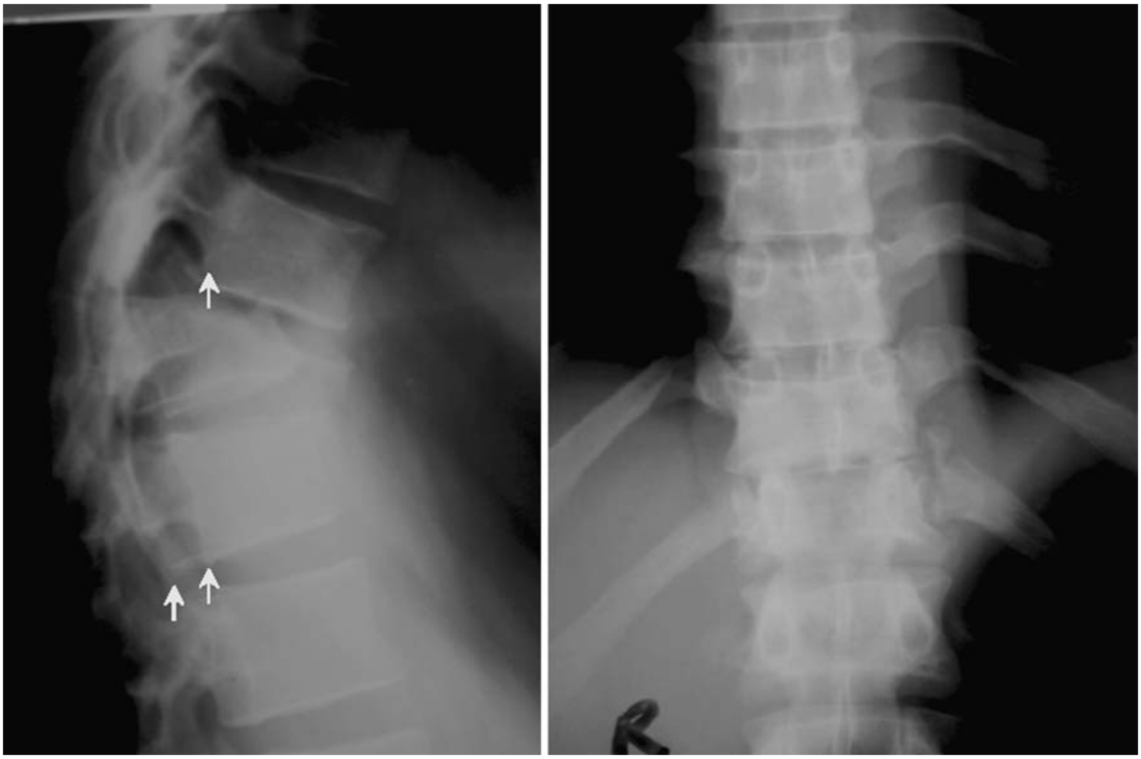

Fig. 4 A collapsing wall fell onto the back of this 35-year-old male. He also sustained an injury at the thoraco-lumbar junction with complete neurological deficit. *Double arrows* on the lateral radiograph indicate the "double line sign", which denote the presence of vertebral rotation when compared to the single arrow cephalad. The two junior residents classified the injury as an AO type B and the remainder of the observers as a type C at the second sitting

integral part of this system. The orthopaedic surgeon should not merely be a technician but a clinician who assimilates and integrates all available data to reach a final diagnosis.

In our opinion, the AO-classification is a simple and comprehensive system, which is helpful in the management of thoraco-lumbar fractures. Although levels of reproducibility, as with most classification systems, is still problematic, improved levels of agreement can be obtained with the integration of both clinical and radiological information.

Acknowledgements The authors thank Ms T.M. Esterhuizen of the Medical Bioethics Unit at the University of KwaZulu Natal, for her kind assistance in performing the statistical analysis.

References

1. Andersen DJ, Blair WF, Steyers CM, Adams BD, el-Khouri GY, Brandser EA (1996) Classification of distal radius fractures: an analysis of interobserver reliability and intraobserver reproducibility. J Hand Surg Am 21:574–582
2. Andersen E, Jorgensen LG, Hededam LT (1990) Evans' classification of trochanteric fractures: an assessment of the interobserver and intraobserver reliability. Injury 21:377–378
3. Blauth M. Bastian L, Knop C, Lange U, Tusch G (1999) Interobserverreliabilitat bei der Klassifikation von thorakolumbalen Wirbelsaulenverletzungen. Orthopade 28:662–681
4. Cohen JA (1960) A coefficient of agreement for nominal scales. Educat Psych Measure 20:37–46
5. Cohen J (1968) Weighted kappa. Nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 70:213–220
6. Denis F (1983) The three column spine and its significance in the classification of acute thoracolumbar spinal injuries. Spine 8:817–831
7. Frandsen PA, Andersen E, Madsen F, Skjodt T (1988) Garden's classification of femoral neck fractures. An assessment of inter-observer variation. J Bone Joint Surg 70-B(4):588–590

8. Frankel HL, Hancock DO, Hyslop G et al (1969) The value of postural reduction in the initial management of closed injuries of the spine with paraplegia and tetraplegia. Paraplegia 7:179–192
9. Gehrchen PM, Nielsen JO, Olesen B (1993) Poor reproducibility of Evans' classification of the trochanteric fracture. Assessment of 4 observers in 52 cases. Acta Orthop Scand 64:71–72
10. Holdsworth FW (1963) Fractures, dislocations and fracture-dislocations of the spine. J Bone Jt Surg Br 45B:6–20
11. Horn BD, Rettig ME (1993) Interobserver reliability in the Gustilo and Anderson classification of open fractures. J Orthop Trauma 7:357–360
12. Kirkwood BR, Sterne JAC (2003) Essential medical statistics, 2nd edn. Blackwell Science, Amsterdam
13. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–174
14. Magerl F, Aebi M, Gertzbein SD, Harms J, Nazarian S (1994) A comprehensive classification of thoracic and lumbar injuries. Eur Spine J 3:184–201
15. Martin JS, Marsh JL, Bonar SK, DeCoster TA, Found EM, Brandser EA (1997) Assesment of the AO/ASIF fracture classification for the distal tibia. J Orthop Trauma 11:477–483
16. Nielsen JO, Dons-Jensen H, Sorensen HT (1990) Lauge-Hansen classification of malleolar fractures. An assessment of the reproducibility in 118 cases. Acta Orthop Scand 61:385–387
17. Oner FC, Ramos LMP, Simmermacher RKJ, Kingma PTD, Diekerhof CH, Dhert WJA, Verbout AJ (2002) Classification of thoracic and lumbar spine fractures: problems of reproducibility a study of 53 patients using CT and MRI. Eur Spine J 11:235–245
18. Swiontkowski MF, Sands AK, Agel J, Diab M, Schwappach JR, Kreder HJ (1997) Interobserver variation in the AO/OTA fracture classification system for pilon fractures: Is there a problem? J Orthop Trauma 2:467–470
19. Thomsen NO, Overgaard S, Olsen LH, Hansen H, Nielsen ST (1991) Observer variation in the radiographic classification of ankle fractures. J Bone Jt Surg Br 73-B:676–678